

An Introduction to Epidemiology

Wei Liu, MPH

Biostatistics Core

Pennington Biomedical Research Center

Baton Rouge, LA

Last edited: January, 14th, 2014

TABLE OF CONTENTS

Introduction.....	1
Terminology.....	2
Study Designs.....	7
References.....	9

Introduction

Epidemiology is the study of the distribution and determinants of diseases or other health-related states or events in specified populations and the application of this study to control health problems (World Health Organization, 2014). As the basic science of public health, Epidemiology is a highly quantitative discipline based on principles of statistics and research methodologies. Epidemiologists study the distribution of frequencies and patterns of health events within groups in a population; and conduct the research to detect the causes or factors that are associated with increased risk or probability of disease.

The objectives of Epidemiology are to:

- identify etiology of disease
- determine the burden of disease
- study the natural history of disease
- evaluate health systems
- provide foundation for public policy

Terminology

1. Incidence

- A measure of new health- or disease-related events occurring in a population, stated as a rate.
- Example: If, over a one-year span, ten people were diagnosed with lung cancer, out of a total study population of 10,000 (who do not have lung cancer at the beginning of the study period), then the incidence of lung cancer in this population was 1/1,000/year.

2. Prevalence

- A measure of an existing outcome at a specific moment in time.
- Example: If a skin cancer screening was taken in a population of 50,000 on July 1st, 2013 and 150 were recently diagnosed with the skin cancer and 350 were living with the skin cancer, then the prevalence of the skin cancer in the population on July 1st, 2013 was 1% ($1\% = (150 + 350) / 50,000$).

3. Ratio

- An expression of the relative frequency of the occurrence of some event compared to some other event.
- Example: In a study group, there were 4 men and 8 women. The ratio of men to women was 1:2 ($4:8 = 1:2$).

4. Odds

- The ratio of the probability of the event of interest to that of the nonevent.
- Example: In a study group, there were 2 people developed diabetes among 100 participants. The odds of developing diabetes was 1:49 ($2 : (100-2) = 2:98 = 1:49$).

5. Odds ratio (OR)

- A measure of association- the odds of exposure for cases divided by the odds of exposure for controls. OR is calculated primarily from case-control studies.
OR < 1, the exposed factor associates with lower odds of outcome;
OR = 1, the exposed factor does not affect odds of outcome;
OR > 1, the exposed factor associates higher odds of outcome.
- Calculating OR:

	Case	Control
Exposure	A	B
No exposure	C	D

Where

A = Number of exposed cases

B = Number of exposed controls

C = Number of unexposed cases

D = Number of unexposed controls

$$OR = \frac{\text{Number of exposed cases/Number of unexposed cases}}{\text{Number of exposed controls/Number of unexposed controls}} = \frac{A/C}{B/D} = \frac{A * D}{B * C}$$

- Example: In a case-control study, researchers recruited 450 participants. The following oral contraceptive use questionnaire found: There were 200 breast cancer participants, which included 140 oral contraceptive users and 60 non-oral contraceptive users. Among the rest 250 participants, 130 oral contraceptive users and 120 non-oral contraceptive users. The odds ratio was 2.15.

	Breast cancer	No breast cancer
Oral contraceptive use	140	130
No oral contraceptive use	60	120

$$OR = \frac{140/60}{130/120} = \frac{140 * 120}{60 * 130} = 2.15$$

6. Relative risk (RR)

- A measure of association- the ratio of the risk (incidence) in exposed individuals divided by that in unexposed. RR is calculated primarily from cohort studies.
RR < 1, the exposed factor decreases the risk of outcome (e.g. disease);
RR = 1, the exposed factor does not change the risk of outcome;
RR > 1, the exposed factor increases the risk of outcome.
- Calculating RR:

	Disease	No disease
Exposure	A	B
No exposure	C	D

Where

- A = Number of exposed people who developed disease
- B = Number of exposed people who did not develop disease
- C = Number of unexposed people who developed disease
- D = Number of unexposed people who did not develop disease

$$RR = \frac{\text{Incidence of disease in exposed group}}{\text{Incidence of disease in unexposed group}} = \frac{A/(A + B)}{C/(C + D)}$$

- Example: In a cohort study, researchers recruited 4,000 participants, 2,000 of them were smoker and 2,000 of them were not. After 10 years follow-up period, 60 people developed coronary heart disease (CHD) and 1,940 people did not in the smoking group. 30 people developed CHD and 1,970 did not in the non-smoking group. The relative risk was 2.0.

	CHD	No CHD
Smoking	60	1940
Non-smoking	30	1970

$$RR = \frac{60/(60 + 1940)}{30/(30 + 1970)} = \frac{60/2000}{30/2000} = 2.0$$

7. Attributable risk (AR)

- An absolute measure of the effect of exposure compared to non-exposure; the excess occurrence of disease in the study sample due to risk factor.
- Calculating AR:
Followed the table for relative risk calculation,

$$AR = \frac{A}{A + B} - \frac{B}{C + D}$$

- Example: Continued the scenario for relative risk, the attributable risk was 15/1,000/year.

$$AR = \frac{60}{60 + 1940} - \frac{30}{30 + 1970} = 0.03 - 0.015 = 0.015 = 15/1000/year$$

8. Attributable risk percent (AR%)

- A measure of the proportion of disease in exposed group that is due to exposure in the study sample.
- Calculating AR%:
Followed the table for relative risk calculation,

$$AR\% = \frac{A/(A + B) - C/(C + D)}{A/(A + B)}$$

- Example: Continued the scenario for relative risk, the attributable risk percent was 50.0%.

$$AR\% = \frac{A/(A + B) - C/(C + D)}{A/(A + B)} = \frac{60/(60 + 1940) - 30/(30 + 1170)}{60/(60 + 1940)} = \frac{0.03 - 0.015}{0.03} = 50.0\%$$

9. Validity

- The ability of a test to distinguish between who have the disease (or other characteristic) and who do not.

Sensitivity

- The ability of a test to identify correctly those who have the disease (or characteristic) of interest.

Specificity

- The ability of a test to identify correctly those who do not have the disease (or characteristic) of interest.
- Calculating sensitivity and specificity:

	Diseased	Not Diseased
Test +	A	B
Test -	C	D

Where

A = Number of true positives (positive results in diseased people)

B = Number of false positives (positive results in non-diseased people)

C = Number of false negatives (negative results in diseased people)

D = Number of true negatives (negative results in non-diseased people)

$$\text{Sensitivity} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{Number of false negatives}} = \frac{A}{A + C}$$

$$\text{Specificity} = \frac{\text{Number of true negatives}}{\text{Number of true negatives} + \text{Number of false positives}} = \frac{D}{D + B}$$

- Example: In a breast cancer study, 1,040 participants went through screening tests. Positive screening result indicates the existence of disease and negative screening result indicates the non-existence of disease. Later the biopsy (gold standard for diagnosis) showed 150 people with breast cancer had positive screening test results; 50 people with breast cancer had negative screening test results; 40 people without breast cancer had positive screening test results; and 800 people without breast cancer had negative screening test results. The sensitivity of the screening test was 75.0% and the specificity of the screening test was 95.2%.

	Diseased	Not Diseased
Test +	150	40
Test -	50	800

$$\text{Sensitivity} = \frac{150}{150 + 50} = \frac{150}{200} = 75.0\%$$

$$\text{Specificity} = \frac{800}{800 + 40} = \frac{800}{840} = 95.2\%$$

10. Reliability (repeatability)

- The extent to which the results obtained by a test are replicated if the test is repeated.

11. Type I error (α)

- The incorrect rejection of a true null hypothesis (H_0). Power and sample size calculations usually set α at 0.05, which means, 5% of the time, we reject the null hypothesis when it is true.

Type II error (β)

- The failure to reject a false null hypothesis. The power equals $1-\beta$. Power and sample size calculations usually set power at 0.8, which means 80% of the time, we reject the null hypothesis when it is false.
- Two types of errors

	True H_0	False H_0
Accept H_0	True	Type II error
Reject H_0	Type I error	True

12. Confounding

- The situation in which a noncausal association between a given exposure and an outcome is observed as a result of the influence of a third variable or group of variables.
- Example: In a cohort study to examine the relationship between alcohol consumption and heart disease, researchers found a third variable-smoking, was associated with alcohol consumption. People who smoke were more likely to consume alcohol. Also, smoking was an independent risk factor for heart disease. Thus smoking was a confounding variable.

13. Bias

- The result of a systematic error in the design or conduct of a study.
- There are two basic classifications of bias: Selection bias and information bias.

Selection bias

- The bias is due to different probabilities of individuals being included in the study sample according to relevant study characteristics (the exposure and the outcome of interest);
- Example: In a case control study to detect relationship between oral contraceptives (OC) use and breast cancer, more breast cancer participants with known OC use history were recruited in the case group. This was the selection bias.

Information bias

- The bias is present when there is a systematic tendency for individuals selected for inclusion in the study to be erroneously placed in different exposure/outcome categories, leading to misclassification.
- Example: In a case control study to detect relationship between oral contraceptives (OC) use and breast cancer, an interview for OC use history was conducted for all participants. Interviewer who knew the case/control status asked breast cancer participants more thoroughly for the OC use history, which lead information bias.

Study Designs

Case reports or case series

- These describe socio-demographic, behavioral and/or medical characteristics for one or more persons with similar diagnosis.
- Example: characteristics of children admitted to a hospital with cerebral malaria during a two-year period.

Ecological studies

- These use populations or groups of individuals as units of observation. The units of observation are usually geographically defined populations or the same geographically defined population at different points in time.
- Exposure to outcome cannot be linked in a given individual.
- An ecologic association might accurately reflect a causal connection between a suspected risk factor and a disease.
- Example: the increase over time in the number of persons working as gem miners along the Thai-Cambodian border, as an exposure, parallels the rise in Plasmodium falciparum malaria cases during the same time period, as an outcome.

Cross-sectional studies

- These examine the relationship between a disease or other health-related characteristic and other variables of interest they exist in a population at a given point in time.
- No information on the temporal sequence of cause and effect can be provided.
- Association is measured via odds ratio.
- Example: surveys to describe characteristics or behaviors within a study population (flu prevalence, vaccine coverage) and / or examine potential risk factors.

Case-control studies

- These examine or test the relationship between specific determinant(s) or exposures and case status.
- Studies start with groups affected with the outcome and groups not affected and retrospectively determine the rates of exposure to a risk factor for each group to see if these rates differed from outcome to exposure.
- Association is measured via odds ratio.
- Example: To assess the relationship between high-dose oral contraceptives and breast cancer, researchers selected the diseased and non-diseased, and then sent our questionnaire for collecting their oral contraceptives usage history.

Cohort studies

- These investigate whether the incidence of an event is related to a suspected exposure.
- Studies start with exposed and unexposed groups and following them to see if the rates of occurrence of the outcome in the two groups differ- from exposure to outcome.
- Association is measured via relative risk.
- Example: To study risk factors for heart disease, researchers measured lifestyle and health status variables every year or two by questionnaire and/or examination, for more than 2 decades, in an adult male population in Framingham, MA.

Clinical trials

- These (in phase 3) compares new therapy to placebo or existing option(s).
- The treatment or exposure is randomly assigned to study subjects by the investigator.
- Example: To find out if a new drug is more effective at preventing death or severe disability in stroke victims than the current standard treatment, researchers randomized 100 incident stroke patients to standard or new therapy and followed for 1 year.

References:

1. World health organization, Introduction to basic epidemiology and principles of statistics for tropical diseases control Part I. Learner's Guide (Accessed Jan 14, 2014) http://whqlibdoc.who.int/hq/2000/WHO_CDS_CPE_SMT_2000.2_Rev.1_Part1.pdf
2. Szklo, M., & Nieto, F. J. (2004). *Epidemiology: Beyond the basics* (2nd ed.). Sudbury, Mass: Jones and Bartlett.
3. Gordis, L. (2009). *Epidemiology* (4th ed.). Philadelphia: Elsevier/Saunders.
4. Greenberg, R. S., Daniels, S. R., Flanders, W. D., Eley, J. W., & Boring, J. R. (1996). *Medical Epidemiology* (2nd ed.). Stamford, Ct: Appleton & Lange.
5. American centers for disease control and prevention, An introduction to Epidemiology (Accessed Jan 14, 2014) http://www.cdc.gov/excite/classroom/intro_epi.htm